

PROCESAMIENTO DE IMÁGENES AÉREAS PARA DETECCIÓN Y MONITORIZACIÓN EMPLEANDO VEHÍCULOS AÉREOS

Luis Merino

Dpto. Ingeniería de Sistemas y Automática. Universidad de Sevilla. e-mail:merino@cartuja.us.es

Aníbal Ollero

Dpto. Ingeniería de Sistemas y Automática. Universidad de Sevilla. e-mail:aollero@cartuja.us.es

Resumen

El presente artículo presenta técnicas de procesamiento de imágenes aéreas para su empleo en tareas de detección y monitorización. En particular, se presentan técnicas para la estabilización de imágenes y geo-localización de características basadas en algoritmos de seguimiento de regiones y modelos del terreno. Dichas técnicas están siendo desarrolladas en el marco del proyecto COMETS del V Programa Marco de la Comisión Europea, dedicado a la coordinación y control en tiempo real de múltiples vehículos aéreos heterogéneos.

Palabras clave: Robótica aérea, UAVs, procesamiento de imágenes aéreas.

1. INTRODUCCIÓN

El uso de secuencias de vídeo tiene un gran interés para aplicaciones de detección y de monitorización, pues permiten observar comportamientos dinámicos en una escena. En labores de monitorización, es habitual obtener dichas secuencias mediante cámaras fijas. Sin embargo, el uso de medios aéreos (helicópteros, aviones) es una excelente alternativa para la disposición de cámaras para labores de detección y monitorización. Al disponerse la cámara sobre un vehículo, es posible controlar la posición de este para situar la cámara de forma adecuada. Además, al tratarse de un vehículo aéreo, es posible monitorizar áreas mayores, sin las limitaciones impuestas por vías de comunicación o irregularidades del terreno, alcanzando zonas de difícil acceso.

En los últimos años, el uso de medios aéreos en tareas automáticas de monitorización y reconocimiento se ha desvelado como una solución factible [11]. Sin embargo, el uso de helicópteros o aviones tripulados conlleva grandes costes y en algunos casos comporta un riesgo importante para los pilotos.

Una alternativa es el uso de vehículos aéreos no tripulados (*Unmanned Aerial Vehicles* o UAV en su denominación inglesa). Durante la última década, el desarrollo de UAV ha ido alcanzando su madurez. Los UAV han ido incrementando su capacidad de procesamiento de información hasta convertirse en robots aéreos dotados de una notable autonomía. En la actualidad existen numerosos proyectos relacionados con ellos, tanto militares como universitarios, fundamentalmente en los Estados Unidos. Así, las universidades de Southern California [5], Carnegie Mellon [1], Georgia Tech [4] tienen o han tenido en marcha proyectos importantes relacionados con la robótica aérea. En Europa, el proyecto WITAS [9] de la Universidad de Linköping en Suecia es uno de los más importantes. La Universidad Técnica de Berlín también ha desarrollado un vehículo autónomo, el MARVIN, que ha demostrado su capacidad para localizar automáticamente objetos [13]. A nivel nacional, pueden mencionarse las actividades del Instituto Nacional de Técnica Aeroespacial (INTA), con su proyecto SIVA, o el proyecto ELEVA de la Universidad Politécnica de Madrid [2].

El pasado mes de mayo se ha iniciado el proyecto COMETS, del V Programa Marco. Este proyecto considera la cooperación de varios UAV heterogéneos (helicópteros y dirigibles) para tareas de detección y monitorización. La posibilidad de coordinar y combinar distintas vistas aéreas de un mismo suceso tiene un gran interés en labores como la detección y localización. Así, el uso de varias vistas de una escena permite confirmar de forma más precisa alarmas en tareas de detección, así como obtener medidas más precisas en labores de monitorización.

El presente artículo muestra técnicas para la estabilización y localización de imágenes tomadas desde vehículos aéreos. también se muestra como combinar imágenes tomadas desde diferentes vehículos aéreos. La sección 2 muestra una introducción a las técnicas básicas. La sección 3 presenta un mecanismo para la eliminación de vibra-

ciones en imágenes. La sección 4 explica como es posible localizar objetos en coordenadas geográficas para escenas planas. La sección 5 muestra un método para la combinación de imágenes tomadas desde distintos puntos de vista, así como resultados. Por último, la sección 6 presenta conclusiones y líneas futuras de desarrollo.

2. TÉCNICAS BÁSICAS PARA EL PROCESAMIENTO DE IMÁGENES AÉREAS

El análisis y la estimación del movimiento en el plano de la imagen constituye una de las operaciones básicas en el procesamiento de secuencias de imágenes. Los resultados de este primer nivel de procesamiento servirán de entrada al procesamiento de un nivel más alto, encargado de tareas específicas tales como la detección automática.

Las técnicas para la estimación del desplazamiento en secuencias de imágenes se pueden dividir fundamentalmente en dos grandes grupos: métodos de flujo óptico [3], que tratan de obtener un campo denso de velocidades en el plano de la imagen; y métodos de seguimiento de regiones, que tratan de obtener los desplazamientos que experimentan un determinado número de regiones a lo largo de la secuencia de imágenes [8].

Ambos métodos presentan ciertas ventajas respecto del otro: así, mientras los métodos de flujo óptico son más adecuados para la detección de objetos en movimiento, los métodos de seguimiento de regiones permiten un mayor desplazamiento entre imágenes. Estos últimos son también más adecuados para el seguimiento de objetos en la imagen.

En el caso de imágenes tomadas desde medios aéreos la cámara está en movimiento. En este caso, el movimiento calculado en la imagen constará de dos componentes: el desplazamiento en los niveles de gris de la imagen inducido por el propio movimiento de la imagen y el desplazamiento debido a objetos con movimiento independiente. La segmentación de ambos tipos de movimiento constituye otro problema importante en el análisis de las imágenes aéreas. Existen distintas técnicas para ello. En general, éstas se basan en emplear modelos de movimiento, relacionados con la configuración de la escena que se está observando y con el modelo de la cámara [6].

El presente artículo se centra en el empleo de técnicas basadas en el seguimiento de regiones y modelos del terreno para la estabilización de las imágenes, así como la geo-localización de los puntos detectados en una imagen.

3. ESTABILIZACIÓN DE LAS IMÁGENES

En el caso de monitorización empleando imágenes aéreas, es frecuente que las cámaras se ven afectadas por vibraciones. Estas vibraciones pueden amortiguarse de forma importante empleando dispositivos estabilizadores tales como gimbals, que tienen un elevado coste y cuyo peso no suele ser admisible para su empleo en vehículos aéreos autónomos, que normalmente tienen una capacidad de carga muy reducida.

Por tanto interesa desarrollar técnicas que permitan la estabilización mediante procesamiento de imágenes.

Si se supone que las vibraciones son rotaciones en torno al centro óptico, y para una escena estática, es posible tomar una imagen como referencia y referir todas las demás a dicha imagen mediante homografías ([14] [7]). Empleando coordenadas homogéneas, dicha homografía queda definida por la ecuación (1).

$$\begin{pmatrix} u_{\lambda,ref} \\ v_{\lambda,ref} \\ \lambda \end{pmatrix} = \underbrace{\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{pmatrix}}_{\mathbf{H}} \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} u_{ref} \\ v_{ref} \end{pmatrix} = \frac{1}{\lambda} \begin{pmatrix} u_{\lambda,ref} \\ v_{\lambda,ref} \end{pmatrix}. \quad (2)$$

donde $\mathbf{x}_i = (u_i, v_i)^T$ es la posición de un píxel en la imagen i , mientras que $\mathbf{x}_{ref} = (u_{ref}, v_{ref})^T$ es la posición del píxel correspondiente en la imagen de referencia. Nótese que la homografía \mathbf{H} está definida salvo un factor de escala. De ahí que uno de los coeficientes pueda tomarse arbitrariamente igual a 1.

Para la obtención de la homografía que relaciona dos imágenes (bajo la suposición de una rotación) bastaría con conocer la correspondencia de 4 puntos (8 coeficientes). Sin embargo, en general, se obtendrán más correspondencias, obteniendo los parámetros de la transformación empleando mínimos cuadrados.

3.1. SELECCIÓN Y SEGUIMIENTO DE REGIONES

Por tanto, para la eliminación de las vibraciones, se seleccionan una serie de puntos en la imagen tomada como referencia. Dichos puntos son seguidos en el plano de la imagen durante la secuencia de imágenes, calculando en cada caso la homografía indicada en las ecuaciones (1) y (2) mediante mínimos cuadrados.

En realidad, lo que se sigue en el plano de la imagen son regiones en torno a cada uno de los puntos (regiones 7×7 en general). Dichas regiones R se buscan en la nueva imagen I_i usando un método basado en la suma de las diferencias al cuadrado (criterio SSD en sus siglas inglesas), es decir, buscando la posición que minimiza $\delta(x, y)$ en:

$$\delta(x, y) = \sum_R (R(k, l) - I_i(x + k, y + l))^2 \quad (3)$$

Es posible obtener una resolución sub-píxel en la búsqueda, ajustando por ejemplo un polinomio cuadrático al valor de δ en un entorno del mínimo inicial encontrado, y usando la posición del mínimo del polinomio como valor final.

En cuanto a los puntos a seleccionar en la imagen de referencia, es importante que las regiones a seguir contengan información suficiente para permitir calcular el desplazamiento que han sufrido, es decir, no sufran el denominado problema de la apertura. Es necesario encontrar regiones que presenten elevado contraste.

Existen distintas medidas acerca de la estructura de una región. Así, es posible emplear medidas basadas en derivadas de la imagen. En [12] y [3], se muestra que, para el criterio SSD, las regiones en las que los autovalores de la matriz:

$$\begin{aligned} & \sum_R W(\mathbf{x}) \nabla I(\mathbf{x}) \nabla I(\mathbf{x})^T = \\ = & \underbrace{\begin{bmatrix} \sum_R W(\mathbf{x}) I_x^2(\mathbf{x}) & \sum_R W(\mathbf{x}) I_x(\mathbf{x}) I_y(\mathbf{x}) \\ \sum_R W(\mathbf{x}) I_x(\mathbf{x}) I_y(\mathbf{x}) & \sum_R W(\mathbf{x}) I_y^2(\mathbf{x}) \end{bmatrix}}_{\mathbf{T}} \end{aligned} \quad (4)$$

son altos son regiones adecuadas (I_x e I_y indican las derivadas parciales de la imagen respecto de las coordenadas x y y respectivamente). En este caso R es la región considerada, y $W(\mathbf{x})$ son pesos asociados a cada píxel de la región (en general, W es igual a 1. También es común emplear una función gaussiana).

En [10] se propone la siguiente medida para el desarrollo de un detector de esquinas:

$$h = \det \mathbf{T} - k \cdot \text{traza}^2(\mathbf{T}) \quad (5)$$

donde $k = 0,04$. Buscando máximos locales sobre el valor de h se pueden determinar regiones que en general no presentan el problema de la apertura.

En la figura 1 se muestran 3 imágenes tomadas desde un vehículo aéreo. También se muestran

algunas de las regiones empleadas para la estabilización, y como estas regiones son seguidas en 3 imágenes consecutivas. Por último, se muestran las imágenes estabilizadas.

3.2. DETECCIÓN DE PUNTOS CON MOVIMIENTO INDEPENDIENTE

La técnica supone que todas las regiones empleadas para determinar \mathbf{H} pertenecen a puntos estáticos de la escena. Sin embargo, es posible que algunas de las regiones seleccionadas pertenezcan a objetos con movimiento independiente (es decir, regiones cuyo movimiento es debido al movimiento del objeto al que pertenecen en el espacio real, ver sección 2).

Para detectar esas regiones puede emplearse el procedimiento que se indica a continuación.

En primer lugar, dado que los puntos correspondientes a objetos con movimiento independiente en general no verifican la ecuación (1), una vez determinada \mathbf{H} , se calculan los residuos:

$$r_k = \|\mathbf{x}_k - \hat{\mathbf{x}}_k\|^{\frac{1}{2}} \quad (6)$$

donde \mathbf{x}_k es la posición de la región y $\hat{\mathbf{x}}_k$ es la posición prevista según las ecuaciones (1) y (2). Si el máximo residuo es mayor que un umbral, la región correspondiente se identifica como perteneciente a un objeto con movimiento independiente y se descarta para el cálculo de \mathbf{H} .

De este modo, sin embargo, sólo es posible determinar cada vez una región (pues dicha región ya ha afectado al cálculo de \mathbf{H}). Para determinar más regiones, se desarrolla un proceso iterativo en el cálculo de \mathbf{H} , eliminando en cada paso una región y volviendo a calcular \mathbf{H} , hasta que el error máximo esté por debajo de un umbral.

Las regiones descartadas para el cálculo de \mathbf{H} son mantenidas, sin embargo, en el procedimiento de seguimiento de regiones, pues en otro caso, el proceso automático de selección volvería a escogerlas.

4. GEO-LOCALIZACIÓN AUTOMÁTICA

Una característica importante que debe proveer un sistema de detección o monitorización automática es la posibilidad de expresar la posición en coordenadas geográficas de un objeto de interés en la imagen, es decir, geo-localizar el objeto.

El presente apartado describe un método para la determinación de la posición de un punto, conociendo su posición en el plano de la imagen.



Figura 1: Estabilización. Imágenes originales e imágenes estabilizadas.

4.1. LOCALIZACIÓN EN ESCENAS PLANAS

Si se supone una superficie plana, la relación entre los puntos en el plano de la imagen y los puntos sobre el terreno vuelve a ser una homografía. La ecuación (1) sigue siendo válida, salvo que ahora (u_{ref}, v_{ref}) hacen referencia a coordenadas sobre el terreno. Por otro lado, dicha homografía está relacionada con la posición y orientación de la cámara, así como su distancia focal. Si se supone un modelo *pin-hole* de la cámara y que, sin pérdida de generalidad, la superficie considerada es el plano $z_w = 0$, se tiene:

$$\begin{pmatrix} x_i \lambda_i \\ y_i \lambda_i \\ \lambda_i \end{pmatrix} = \begin{pmatrix} u_1 & u_2 & -\hat{\mathbf{u}}^T \mathbf{p} \\ r_1 & r_2 & -\hat{\mathbf{r}}^T \mathbf{p} \\ n_1/f & n_2/f & -\hat{\mathbf{n}}^T \mathbf{p}/f \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ 1 \end{pmatrix} \quad (7)$$

donde \mathbf{p} es el vector de posición de la cámara, $\hat{\mathbf{u}} = (u_1, u_2, u_3)^T$, $\hat{\mathbf{r}} = (r_1, r_2, r_3)^T$, $\hat{\mathbf{n}} = (n_1, n_2, n_3)^T$ es la base ortonormal que define el sistema de coordenadas de la cámara (ver figura 2), $\mathbf{x}_w = (x_w, y_w)^T$ son las coordenadas de un punto en el sistema global y $\mathbf{x}_i = (x_i, y_i)^T$ son las coordenadas de la proyección de ese punto en el plano de la imagen.

Por tanto, si se tiene una estimación inicial de estos parámetros, es posible calcular una estimación inicial de \mathbf{H} . Una vez conocida, es necesario actualizar la perspectiva en cada imagen de la secuencia. Para ello, el procedimiento es el siguiente:

- Mediante la medida de Harris (ecuación (6)) se seleccionan regiones. La posición en el es-

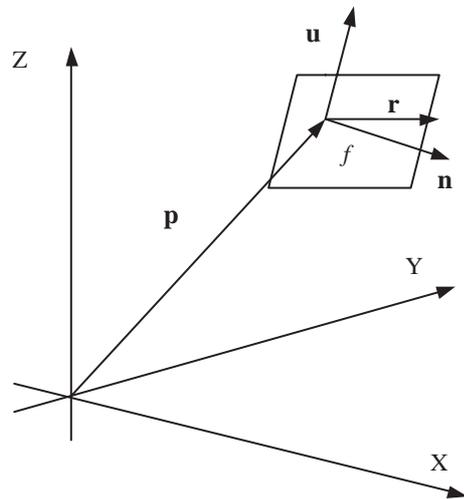


Figura 2: Sistemas de coordenadas.

pacio de dichas regiones se obtiene con la transformación inicial.

- Mediante el método descrito en la sección 3.1 se hace un seguimiento en cada imagen de las regiones, obteniendo las nuevas posiciones $\mathbf{x}_i = (x_i, y_i)^T$ y actualizando \mathbf{H} en cada caso.

También en este caso es necesario detectar aquellas regiones pertenecientes a objetos con movimiento independiente. Para ello se seguirá el mismo procedimiento que el descrito en el apartado anterior.

La utilidad de un modelo plano del terreno depende de la aplicación, pero para escenas lejanas, la mayoría de las superficies pueden ser aproximadas por un plano. Una alternativa más general

consistiría en dividir la escena en zonas planas distintas, y calcular la homografía correspondiente a cada una de las zonas. En caso de tener terrenos más complejos sería necesario considerar un modelo más completo.

5. COMBINACIÓN DE DISTINTAS VISTAS AÉREAS

En el proyecto COMETS, se considera la coordinación de múltiples UAV, que pueden obtener distintas vistas de una misma escena. El combinar las imágenes obtenidas reviste gran interés, al permitir reducir errores, permitiendo además la monitorización con oclusiones temporales de la escena desde algunos UAVs.

Puesto que en el sistema COMETS los distintos UAV pueden observar una escena desde puntos de vista muy alejados entre sí, no es factible tratar de relacionar unas vistas con otras empleando procesamiento sobre los niveles de gris en la imagen, como técnicas de correlación, puesto que lo observado diferirá en gran medida.

Es posible incluso considerar la situación en la que se tienen imágenes de diferentes zonas del espectro, como imágenes infrarrojas y visuales.

Un modo de combinación de las vistas consiste en aplicar la técnica descrita en el apartado 4 a cada una de las vistas. Si las imágenes pueden referirse a unas mismas coordenadas geográficas, es posible establecer relaciones entre ellas sobre el terreno (figura 4).

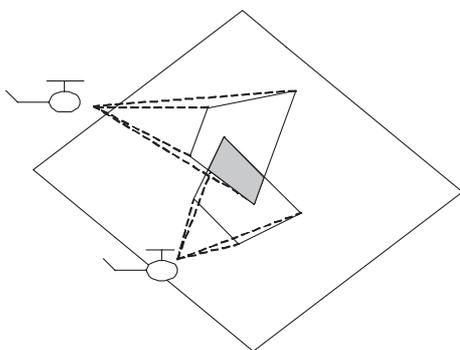


Figura 4: Combinación de diferentes vistas.

La figura 3 muestra el resultado de aplicar lo anterior a un par de imágenes, una terrestre y otra aérea, de la misma zona.

6. CONCLUSIONES

Se han presentado técnicas, basadas en seguimiento de regiones, para la estabilización de imágenes sometidas a vibraciones, así como la geo-

localización automática de los objetos de interés en el plano de la imagen, suponiendo un modelo plano de la escena. Se ha optado por técnicas de seguimiento de regiones, puesto que permiten tratar desplazamientos mayores entre imágenes que otro tipo de técnicas como flujo óptico.

También se ha mostrado como pueden combinarse varias vistas distintas de una misma escena, haciendo uso de las técnicas de geo-localización.

En cuanto a los próximos pasos, cabe mencionar la aplicación de técnicas de estimación de las homografías más robustas (tanto para la estabilización como la localización), como Least Median Squares (LMedS) o M-estimadores [15]. Estas técnicas son más robustas ante datos con errores debidos al seguimiento.

También es posible considerar modelos más complejos del terreno o del movimiento de la cámara para la estabilización y localización.

Asimismo, se pretende profundizar en la combinación de imágenes infrarrojas y visuales de una misma escena, tomadas desde distintos puntos de vista.

Agradecimientos

El trabajo descrito ha sido desarrollado en el seno de los proyectos SPREAD (Comisión Europea, 5º Programa Marco, EVG1-CT-2001-00043) y COMETS (Comisión Europea, 5º Programa Marco, IST-2001-34304). Los autores agradecen la colaboración de Begoña Arrue, Joaquín Ferruz, Francisco Gómez y J.R. Martínez de Dios en el desarrollo del trabajo. Los autores también agradecen la ayuda del Prof. Viegas y el equipo de ADAI de la Universidad de Coimbra (Portugal) en la toma de imágenes aéreas.

Referencias

- [1] O. Amidi, T. Kanade, y K. Fujita. A visual odometer for autonomous helicopter flight. En *Proceedings of IAS-5*, 1998.
- [2] A. Barrientos, J. Cerro, y I. Aguirre. Development of a low cost autonomous minihelicopter for power lines inspections. En *SPIE*, 2000.
- [3] J.L. Barron, D.J. Fleet, y S.S. Beachumin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12:43–77, 1994.
- [4] H.B. Christophersen, M. Dhingra, R. Guily, M.G. Hart, E.N Johnson, y A. Kahn. Development of an autonomous aerial reconnaissance system at Georgia Tech. En *2001 Aerial*



Figura 3: Combinación de diferentes vistas. Se muestran dos vistas de la misma escena. La tercera imagen muestra la combinación de ambas, mostrando cómo se vería la segunda imagen superpuesta sobre la primera.

- Robotics Competition*. School of Aerospace Engineering, 2001.
- [5] A.H. Fagg, M.A. Lewis, J.F. Montgomery, y G.A. Bekey. The USC autonomous flying vehicle: an experiment in real-time behaviour-based control. En *Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1993.
- [6] G. Farnebäck. Spatial Domain Methods for Orientation and Velocity Estimation. Lic. Thesis LiU-Tek-Lic-1999:13, Dept. EE, Linköping University, SE-581 83 Linköping, Suecia, Marzo 1999b. Thesis No. 755, ISBN 91-7219-441-3.
- [7] O. Faugeras y Q-T Luong. *The Geometry of Multiple Images*. The MIT Press, 2001. ISBN 0-262-06220-8.
- [8] J. Ferruz y A. Ollero. Real-time feature matching in image sequences for non-structured environments. applications to vehicle guidance. *Journal of Intelligence and Robotic Systems*, 28:85–123, 2000.
- [9] Gösta Granlund, Klas Nordberg, Johan Wiklund, Patrick Doherty, Erik Skarman, y Erik Sandewall. WITAS: An intelligent autonomous aircraft using active vision. En *Proceedings of the UAV 2000 International Technical Conference and Exhibition*, Paris, Francia, Junio 2000. Euro UVS.
- [10] C.G. Harris y M. Stephens. A combined corner and edge operator. En *Proceedings of the 4th Alvey Vision Conference*, Septiembre 1988.
- [11] R. Kumar, H. Sawhney, S. Samarasekera, S. Hsu, H. Tao, Y. Guo, K. Hana, A. Pope, R. Wildes, D. Hirvonen, M. Hansen, and P. Burt. Aerial video surveillance and exploitation. *Proceedings of the IEEE*, 89(10):1518–1539, Octubre 2001.
- [12] B.D. Lucas y T. Kanade. An iterative registration technique with an application to stereo vision. En *Proc. 7th Intl. Joint Conf. on Artificial Intelligence (IJCAI)*., Agosto 1981.
- [13] M. Musial, U.W. Brandenburg, y G. Hommel. Cooperative autonomous mission planning and execution for the flying robot MARVIN. *Intelligent Autonomous Systems*, 6:636–643, 2000.
- [14] J.G. Semple y G.T. Kneebone. *Algebraic Projective Geometry*. Oxford University Press, 1952. ISBN 0-19-853172-9.
- [15] Z. Zhang. Parameter estimation techniques: a tutorial with application to conic fitting. *Image and Vision Computing*, 15:59–76, 1997.